

Counterfeit Detection for New and Old Currency Designs

Anne P. Hillstrom^{*}, Ira H. Bernstein^{**}

The University of Texas at Arlington, Department of Psychology

ABSTRACT

To test the effectiveness of counterfeit deterrence features recently introduced in US currency, observers were asked to discriminate genuine from counterfeit bills using a two-alternative forced-choice task. In Experiment 1, observers judged \$100s with the new and old designs after receiving training in the deterrence features of each design. The counterfeits were representative of two common print processes: inkjet and offset printing. Judgments were made on whole bills, on individual features with the whole bill unmasked, and on individual features with only that feature visible. In Experiment 2, different observers judged \$100s without any training. They then were trained and judged \$50s and \$20 bills. Taken together, the two experiments indicate that people are good at detecting counterfeits, that inkjet counterfeits are easier to detect than offset counterfeits, and that counterfeits of the newly designed bills are easier to detect than counterfeits of the older series. The design improvement was greatest with \$100 bills and, to a lesser extent, \$50 bills. Improvement was minimal with \$20 bills, very likely because observers were very accurate for both series of \$20s. Finally, some deterrence features were more useful than others in aiding discriminations.

Keywords: evaluation, behavioral, counterfeit, detection

1. INTRODUCTION

When people do not attend to the currency they receive, they are sometimes passed very crude counterfeits. But when people focus on finding counterfeits, they are good at detecting them. Surprisingly few studies have assessed people's ability to detect counterfeits. A US federal research laboratory conducted at least one study for the US Treasury's Bureau of Engraving and Printing (BEP) (1), and Stillitz (as reported in 1) conducted an extensive series of proprietary studies for the Bank of London. Collins et al report that both studies found that when people spent time looking at money, they discriminated genuine from counterfeit bills reasonably well. But fewer counterfeits were detected when the design of money made it easy for people to exchange bills of the proper denomination without looking at bills closely. When specifically looking for counterfeits, there was substantial variability among observers in whether genuine bills were reported as counterfeit or whether counterfeits were reported as genuine. However even when classifying bills inaccurately, people spent more time evaluating counterfeit bills than genuine bills, and when prompted about each bill, generated more comments about irregularities in counterfeits than in genuine bills. So people do discriminate counterfeits from genuine currency.

Both studies concluded that more counterfeits would be detected if designs of different denominations were made more similar, so that people had to look more closely to discriminate denominations. For US currency, two recommendations dominated: (a) make all bills multicolored, in order to take advantage of people's ability to discriminate very small color differences, and (b) move denomination indicators towards the portraits, in order to increase the likelihood that people would inspect the portraits, which provide more cues to counterfeits than other parts of the bills (1). It is interesting, though, that the most effective way to deter counterfeiting also turned out to have a serious drawback: reducing the accurate identification of denominations, which would lessen people's confidence in their currency.

One possible reason research about counterfeit detection is scanty in the US is that the BEP made no major changes to the design of bills until the late 1980s when advances in computer technology changed the methods of counterfeiting. But beginning in 1996, a major change was made in the design of US \$20, \$50, and \$100 bills in order to make counterfeiting more difficult. The design changes did not make various denominations less discriminable. Bills still

^{*} hillstrom@uta.edu; phone 1 817 272-5248; fax 1 817 272-2364; <http://www.uta.edu/psychology/faculty/Hillstrom>; Box 19528, University of Texas, Arlington, TX 76019-0528; ^{**} bernstein@uta.edu; phone 1 817 272-3183; <http://www.uta.edu/psychology/faculty/Bernstein>; all other information is the same for both authors.

have denomination indicators at all corners. But the portraits on bills were made larger and were offset from center so that money folding would have less effect on the quality of portraits on genuine bills. Ink that changes color as the viewing angle is changed was incorporated into one feature of the bills in order to make color reproduction more difficult. A watermark was added into the paper. And a number of other detailed changes were made as well.

The BEP wished to evaluate the success of these changes, and this study presents the results of that evaluation. The basic design compared 1996 (new) and 1990 (old) series of bills, looked at various denominations, examined differences between detection of counterfeits printed using an inkjet printer versus an offset method of printing, and examined individual features of the bills, especially those that had either been introduced in the 1996 series (e.g., color shifting ink) or modified from the previous series (e.g., portraits) to deter counterfeiting.

The research was conducted to evaluate how well counterfeits could be detected under optimal circumstances: on each trial, a genuine and a counterfeit bill were presented, and the observers' task was to decide which one was counterfeit. This is the classical psychophysical *two-alternative forced-choice* method that is used to eliminate bias towards thinking most bills are genuine or most bills are counterfeit. Such bias has already been documented (1).

Detection accuracy for specific features of bills was examined under *masked* viewing conditions in which each feature was presented with the remaining parts of the bill masked off to keep other features from providing cues to examine the effectiveness of individual counterfeit deterrence features. Accuracy in judging these specific features was also examined under *unmasked* viewing conditions to see how well the features were used under more realistic viewing conditions. Finally, *global* judgments were also obtained from unmasked bills, allowing respondents to take into account any or all features they wished.

The basic design also involved evaluating the potential change in effectiveness of the 1996 currency series relative to the older 1990 series, so each series could also be judged in absolute terms. Every factor of interest was varied as completely as possible, with some necessary limitations described below. The project was divided into two experiments.

2. EXPERIMENT 1

In Experiment 1, observers judged \$100 bills after receiving training about their design features. Both offset and inkjet counterfeits were used.

2.1 Method

2.1.1 Participants

Seventy-four observers made judgments in this study. Four were recruited from employees at a local grocery store chain. These observers were paid for their participation. The remaining observers were undergraduates at the University of Texas at Arlington who volunteered in order to receive research credit for a course assignment.

Taken collectively, the sample had 13.3 ± 1.2 years of schooling (range = 12 to 16), an average age of 22.7 ± 6.7 years (range = 18 to 65), and was divided equally by gender. The composition by race was 11.1% African-American, 59.2% White, 15.28% Hispanic, and 11.1% Asian. Answers to other demographic questions about money handling experience are in the Appendix.

2.1.2 Stimuli

As noted above, each trial consisted of the presentation of a pair of bills: one counterfeit and one genuine. The bills were arrayed one above the other with location of the genuine bill randomized. A different pair of bills was used in each trial of the experiment. Counterfeit bills were obtained from the same series (1990 or 1996) as the genuine bill with which they were paired. The counterfeit bills used in the study were selected and loaned by the Secret Service. Because they had been used in criminal investigations they had a variety of markings, so each bill of genuine currency was marked in a comparable manner.

Counterfeiters use a variety of print processes, but only two, inkjet printing and offset printing, were represented in this study because of the limited number of counterfeit bills available created using other processes. The bills were selected by secret service agents with expertise in analysis of counterfeits to represent varying counterfeit quality for each feature. Where possible, offset bills created by the same counterfeiter using the same printing apparatus were used in the 3 judgment conditions (featural-masked, featural-unmasked, and global) to minimize stimulus quality differences between judgment conditions. This level of matching was only possible for bills printed by offset printing methods. Inkjet bills required a more subjective matching process. Both genuine and counterfeit bills ranged greatly in age and wear.

The bills for each trial were encased in a plastic covering. The covering was completely opaque over all but the feature to be judged (masked conditions) or over only the margins of the bill (unmasked conditions). Everywhere else it was transparent. These coverings also prevented observers from handling the currency directly; thus limiting judgments to visual features. The mask contours were separated by approximately 1-3 mm. from the relevant feature. The currency pairs were kept in binders when not in use.

Each observer made judgments in 124 trials, which were divided among five judgment conditions. The featural-masked, featural-unmasked, and global conditions were all used for the 1996 series of bills were, but only the featural-masked and global conditions were used for the 1990 series to minimize possible boredom effects.

2.1.2.1 1996 Series

There were six features of interest in the 1996 series: watermark, portrait, fine lines in the pictures, security thread, color shifting ink, and microprinting (very tiny printing on the bill). In the featural judgments of masked bills (henceforth, condition FM96), two trials (and thus two stimulus pairs) were run for each combination of these six features and two counterfeiting methods for a total of 24 trials. This was also true for featural judgments of unmasked bills (condition FU96). In the global judgments (condition GU96), these same 24 trials were run but, in addition, 12 more trials (6 per counterfeiting process) employed counterfeit bills in which at least one of the counterfeit-deterrent features was missing, providing a total of 36 trials.

2.1.2.2 1990 Series

Only four counterfeit deterrence features were available for evaluation in the 1990 Series bills: fine lines, portrait, security thread, and microprinting. All of these had different designs in the 1996 series. In the featural judgments of masked bills (condition FM90), two trials were run with each of the four features and two counterfeiting methods providing a total of 16 trials. Finally, the global judgments (condition GU90) used these 16 same trials plus eight more trials (four per counterfeiting process) in which the counterfeit bill was missing at least one counterfeit deterrence feature, providing a total of 24 trials.

Two parallel sets of the 124 stimulus pairs were constructed; one of which was randomly assigned to a given observer.

2.1.3 Procedure

Observers were tested in a quiet laboratory room set up like a sparsely furnished office. The observers were seated in the room facing a torchiere floor lamp that provided a bright light that could be used comfortably for backlighting when looking through bills at features within the paper. Other than that, no aids were available for inspecting bills. Two BEP security officers monitored the study for security purposes from an adjacent room through a one-way viewable glass. The observers were informed that officers were observing the experiment.

The session began with the experimenter collecting demographic information from the observer (see Appendix 1). The observer read a brochure called "New Designs for Your Money" published by the U.S. Department of Treasury. The experimenter then pointed out the deterrence features and gave the observer a chance to ask questions. This was followed by 2 practice trials before the experiment itself began.

The observer's task was to use a 6-point rating scale to say which bill was genuine (1 = very sure that the genuine bill was on top, 2 = genuine bill was on the top, 3 = guessing that the genuine bill was on top, 4 = guessing that the genuine bill was on the bottom, 5 = genuine bill was on the bottom, and 6 = very sure the genuine bill was on the bottom). For the analyses reported here, responses 1-3 were pooled into a "top" category and responses 4-6 were pooled into a "bottom" category, and so only accuracy, not confidence, is analyzed.

Trials were blocked by the five judgment conditions. Judgment condition was counterbalanced across observers save that observers either completed the two global judgment conditions (GU96 and GU90) before the three featural-judgment conditions (FM96, FU96, and FM90) or vice versa. In other words, they were not asked to switch from trial to trial between judging whole bills and judging individual features.

2.2 Results

2.2.1 Overall Accuracy As A Function Of Experience

Overall accuracy was 93.8%. Accuracy ranged from 83.1% to 99.2% across subjects, except for one outlier whose accuracy 62.1%. This high level of accuracy caused the data to be negatively skewed. In order to compensate for this, the .01 significance level of significance was used throughout the study.

Perhaps the most important demographic result was that observers who had never heard of the counterfeit detection features or who had heard of them but had not looked at them performed more poorly than observers who had looked at them, regardless of whether they had ever tried to detect counterfeits using these features (91.3% and 91.8% v 96.1% and 95.9%), $F(3, 70) = 4.56$.

Observers were asked at the end of the study whether their judgments were based partially on any features other than the ones on which they were instructed to focus. A total of 41% noted at least one additional feature. Nearly half of these (47%) said they responded on the basis of color. In addition, 25% said they considered print quality, 11% said they responded on the basis of overall look, and 8% said they considered fibers or paper. It is important to note that even when judging masked bills, these features may have been available to influence judgments.

2.2.2 Accuracy As A Function Of The Characteristics Of The Currency

Table 1 contains response accuracies as a function of judgment condition, feature, and print process. Overall accuracy was very good in all five conditions (FM96: 90.0%, FU96: 95.0%, GU96: 98.3%, FM90: 85.4%, and GU90: 94.0%). These accuracies were examined using within-subjects analysis of variance (ANOVA). It was not possible to completely cross all variables because some features were only available in 1996 bills, so no single ANOVA could evaluate the results. Four separate analyses were conducted instead. The first looked at differences within the 1996 series, the second looked at differences within the 1990 series, and the remainder looked at differences between the 1990 and 1996 series.

Preliminary analyses indicated that accuracy of global judgments was unaffected by whether observers had prior practice in doing featural judgments, so order of conditions was ignored, and data to be reported were collapsed across the order manipulation. Also note that even though each analysis compares differences among features, the numbers of features varies. For example, conditions FM96, FU96, and GU96 have six features in common, but conditions FM90 and GU90 only have four features in common. This difference in number of features affects other comparisons, such as those involving series and print processes, since these will involve comparisons averaged over different sets of features.

2.2.2.1 1996 Series

The first ANOVA compared percentage correct, pooling over the two replications, for the three 1996 series judgment conditions (FM96, FU96, and GU96), the six common features common to them (watermark, security thread, portrait, fine lines, color-shifting ink, and microprinting) and print process (offset or inkjet). Experimental error for each effect was the interaction of the effect with observers, i.e., error terms were not pooled. Note that although observers were instructed not to focus on individual features in Condition GU96, the analysis looked at the effect of feature for that

condition, too, because trials were set up using counterfeits that matched those used in Conditions FM96 and FU96. Any effect of feature in Conditions FM96 and FU96 reflects both usefulness of those features and error variance caused by selection of different bills for the different features, but any effect of feature in Condition GU96 primarily reflects error and imprecision in matching. Table 2 summarizes the ANOVA effects.

All main effects and interactions were significant. As shown in Figure 1, overall accuracies in judging watermarks (95.8%), portraits (95.3%), thread (96.5%), and color-shifting ink (95.4%) were similar, but accuracies for fine lines (92.6%) and microprinting (91.1%) were lower. Post-hoc comparisons confirm these featural differences. As shown in Figure 2, observers found it easier to detect inkjet counterfeits than offset ones (97.4% v 91.5%). They found masked/featural judgments (FM96) the most difficult (90.0%), unmasked/global (GU96) easiest (98.3%) with unmasked/featural judgments (FU96) intermediate (95.0%). Post hoc comparisons found that Condition FM96 was significantly less accurate than Condition FU96, and Condition FU96 was significantly less accurate than Condition GU96. Differences between print processes ranged from 11.5% with microprinting and 9.0% with fine lines to 0.1% with security thread. The interaction among all three variables was even more complex. The feature by print process interaction arose because observers were 12.6% more accurate in judging inkjets than offsets in condition FM96 and 5.6% more accurate in condition FU96, but there was no difference for condition GU96. The scale used when differences are plotted is from the minimum to the maximum value. In order to insure that the findings were not unduly influenced by the major difference in task between the featural judgment conditions on the one hand and the global judgment condition on the other, the ANOVA was rerun using only conditions FM96 and FU96. All main effects and interactions remained significant.

2.2.2.2 1990 Series

The ANOVA of the 1990 series data followed the same format as the format of the 1996 series data except that there were only two judgment conditions: FM90 (masked/featural) and GU90 (unmasked/global) with only four features common to both (portrait, fine lines, security thread, and microprinting). Table 2 contains the associated F-ratios. Differences among the four features were small and insignificant, accuracy ranging from 92.9% for fine lines to 87.3% for thread, as shown in Figure 3. Only three effects were significant in this case: the main effects of print process and judgment condition and their interaction. As shown in Figure 4, observers again were more accurate in judging inkjets than offsets (96.0% v 83.4%) and in condition GU90 than in condition FM90 (85.5% v 94.0%). Finally, observers did 19.3% better in judging inkjets than offsets in condition FM90, but did only 5.9% better in condition GU90. Differences among the four features were small and insignificant, ranging from 92.9% for fine lines to 87.3% for thread.

2.2.3 Comparing 1990 And 1996 Series: Masked/Featural Judgments

An ANOVA of the masked/featural judgments looked at the effect of four features (fine lines, thread, color shifting ink, and microprinting), two series (from conditions FM96 and FM90) and two counterfeit print processes (inkjet and offset). These F-ratios are presented in Table 3. Every effect except that of feature, which barely missed the .01 criterion, was significant. Overall accuracy was 95.4% in judging inkjets and 79.1% in judging offsets. Accuracy was 89.0% in judging the 1996 series and 85.5% in judging the 1990 series. The data contributing to these analyses are presented graphically in Figures 5-7. The interaction among the three variables is complex, so some further tests were done.

We also examined whether the method of printing counterfeits was differentially effective for individual features. Observers did 23.0% better in judging inkjet microprinting as opposed to offset microprinting, 17% better at judging inkjet portraits than offset portraits, 16% better at judging inkjet fine lines than offset fine lines, and 9.1% better at judging inkjet thread than offset thread. All of these differences were significant. This is another way of describing the result that counterfeits printed by offset print processes are more difficult to detect than counterfeits printed by inkjet print processes, regardless of which feature the observer is looking at.

Analyses also examined effects of series for each of the features, separately for inkjet and offset processes, using the masked stimuli. In each case, the variables were the 1990 and 1996 series and the four features common to the two series (portrait, fine lines, thread, and microprinting). Performance was uniformly high with the inkjet process, so neither main effect nor the interaction was significant. The trend was towards better performance in judging thread with

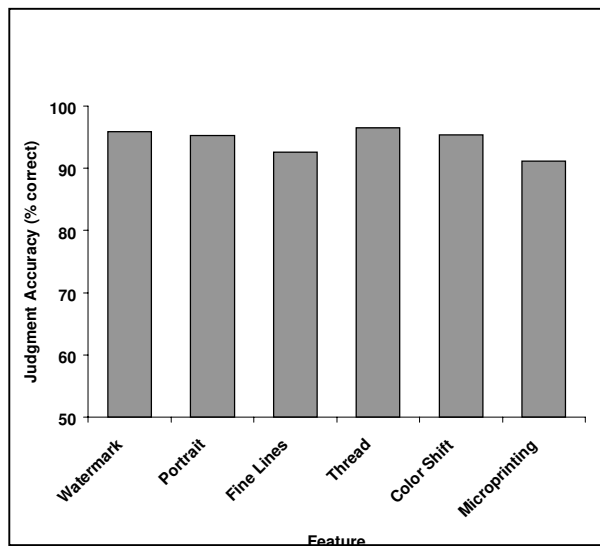


Fig. 1: Effect of Feature on Judgments of 1996 series \$100 currency (conditions FM96, FU96, and GU96 of Experiment 1)

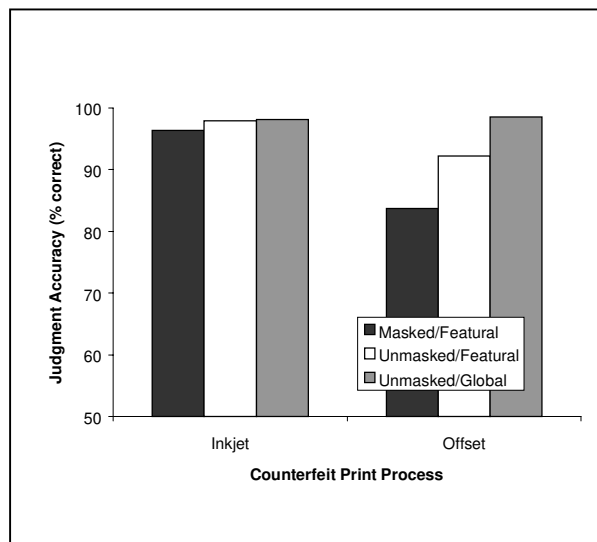


Fig. 2: Effect of Print Process and Judgment Condition of Judgments of 1996 Series \$100 Currency (Conditions FM96, FU96, and GU96 of Experiment 1)

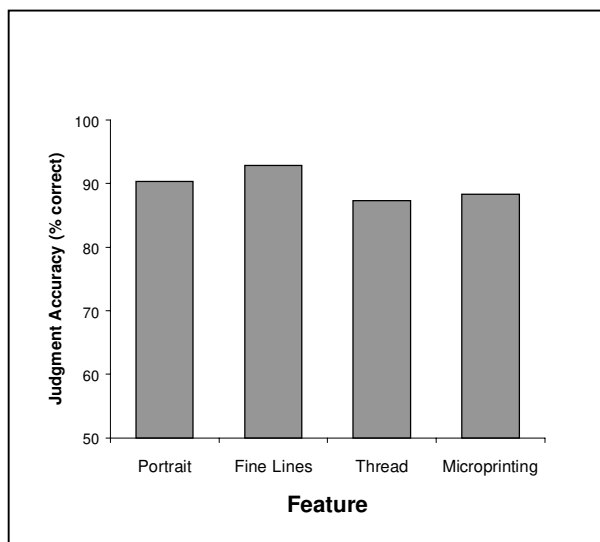


Fig. 3: Effect of Feature on Judgments of 1990 Series \$100 Currency (Conditions FM90 and GU90 of Experiment 1)

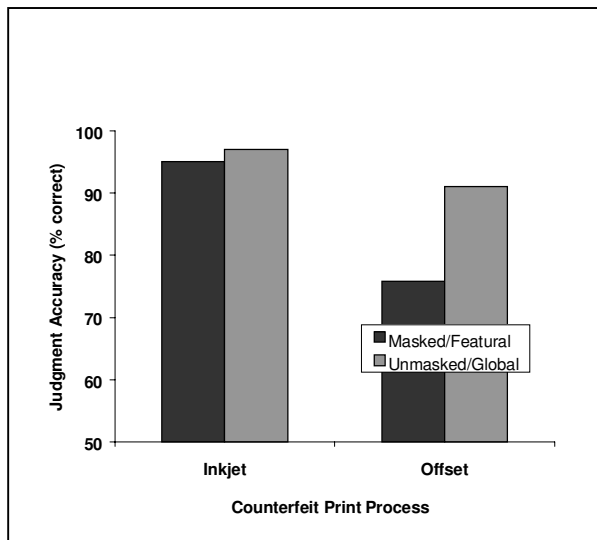


Fig. 4: Effect of Print Process and Judgment Condition on Judgments of 1990 Series \$100 Currency (Conditions FM90 and GU90 of Experiment 1)

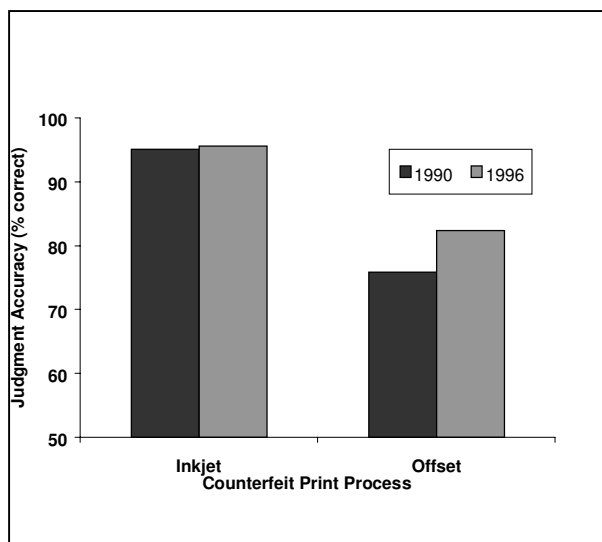


Fig. 5: Effect of Print Process and Series on Featural Judgments of \$100 Currency (Conditions FM96 and FM90 of Experiment 1)

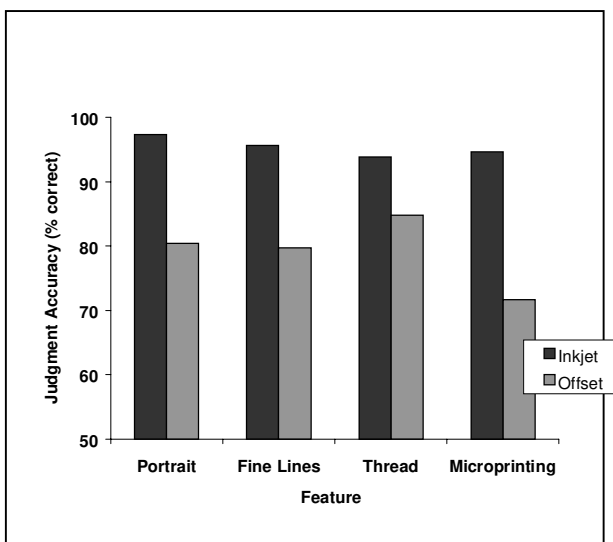


Fig. 6: Effect of Print Process and Feature on Featural Judgments of \$100 Currency (Conditions FM96 and FM90 of Experiment 1)

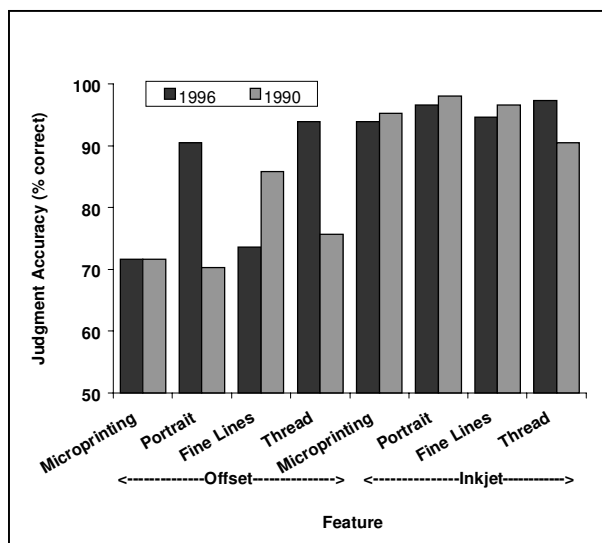


Fig. 7: Masked Featural Judgments of \$100 Currency (Conditions FM96 and FM90 of Experiment 1)

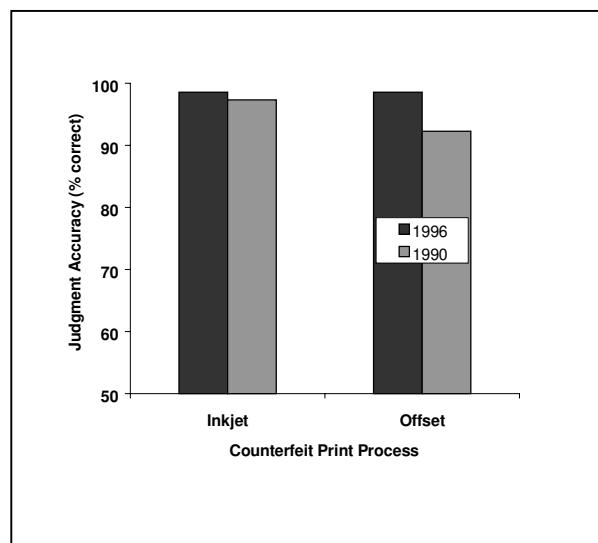


Fig. 8: Effect of Print Process and Series on Global Judgments of \$100 Currency (Conditions GU96 and GU90 of Experiment 1)

new, as opposed to old, series (6.7%)—the remaining three features differed by no more than $\pm 2.5\%$. By contrast, there was a main effect of series for judgments of offset bills, $F(1,73) = 9.35$, a main effect of feature, $F(3,219) = 5.01$, and an interaction, $F(3,219) = 11.20$. Performance was better in the 1996 series when portraits and threads were judged (20.2% and 18.2%), equal when microprinting was judged, but 12.2% worse when fine lines were judged. The latter is the only major instance of poorer performance with the new series bills in the experiment.

2.2.4 Comparing 1990 And 1996 Series: Unmasked/Global Judgments

An ANOVA of the unmasked/global judgments looked at the effect of five levels of feature (fine lines, portrait, thread, microprinting, and feature missing), two series (conditions GU96 and GU90) and the two print processes (offset and inkjet). The F -ratios appear in Table 3, and the response accuracies are presented graphically in Figure 8. Observers correctly judged inkjets 97.9% of the time and offsets 95.5% of the time. Observers judged the 1996 series with an accuracy of 98.6%, and judged the 1990 series with an accuracy of 94.8%. There was only a 1.3% improvement with the 1996 series as compared to the 1990 with inkjets but a 6.4% improvement with offsets. There was not a significant difference between judgments of bills chosen to represent different features.

Source	Judgment Condition					
	Masked/Featural			Unmasked/Global		
	<u>df(Eff.)</u>	<u>df(Err)</u>	<u>F</u>	<u>df(Eff.)</u>	<u>df(Err)</u>	<u>F</u>
Feature (A)	3	219	3.84	4	292	2.95
Process (B)	1	73	193.60*	1	73	11.92*
Series (C)	1	219	8.61*	1	292	19.56*
A*B	3	73	4.28*	4	73	2.82
A*C	3	219	12.07*	4	292	4.74*
B*C	1	73	5.91*	1	73	18.43*
A*B*C	3	219	7.19*	4	292	1.87

Note: df(Eff.) denotes the df for the effect in question, and df(Err.) is the error df. An alpha level of .01 was chosen because of the negative skew in the data.

Table 3: F-ratios for ANOVAs comparing feature, print process, and series, Experiment 1

Because observers were not directed to judge individual features, they presumably made judgments based on the same aspects of the bills, at least within a series of trials. As expected, there was a difference between inkjet and offset bills, reflecting the substantial difference in quality of counterfeits between them. One interpretation of the significant interaction between feature and series is that there is variation in the quality of the bills across the featural conditions, which categorizes by series and print process and feature. It is therefore also possible that offset 1990 series bills chosen to represent portraits may have been of higher quality than bills chosen to represent other features, and that inkjet 1990 series bills chosen to represent security thread may have been of lower quality than bills chosen to represent other features. Broken down at this level, the 1996 bills did not seem to vary as much in quality as the 1990 series bills.

2.3 Discussion of Experiment 1

These data establish clearly the expected finding that judging inkjets is easier than judging offsets. This was the case even though a stringent (.01) criterion for rejecting the null hypothesis was used to make allowance for the inherent skewness in the data. These data also establish the more important point that detecting counterfeit bills was easier with the 1996 series design than the 1990 design. This improvement was clearer with judgments of the offset bills because of the ceiling effect present with inkjet bills. Direct comparisons between judgments of 1990 and 1996 series bills further suggest that judgments about the security thread and the portrait were most improved. Fine lines were judged more poorly with the 1996 series than with the 1990 series. The watermark and color-shifting ink could not be included in this series comparison because they were first introduced in the 1996 series. Looking only at the 1996 series and focusing on

comparisons with offset counterfeits, accuracy in judging these two features was intermediate compared to the other features for the offset counterfeits.

One additional finding of note is that accuracy was greatest in making global judgments of unmasked bills; unmasked/featural judgments and masked/featural judgments were judged less accurately. There are probably a number of reasons for this. For instance, the quality variation in the counterfeits may affect multiple features of a bill in the same way. For example, a relatively unskilled counterfeiter may be both unable to draw a convincing copy of the portrait and the watermark. This could imply synergism (the whole being greater than the sum of its parts), but global judgments may simply give observers multiple opportunities to detect that a bill is counterfeit. If observers could not discriminate the bills based on the first feature they looked at, they could look at another feature in the unmasked/global conditions.

3. EXPERIMENT 2

In Experiment 2, observers judged offset and inkjet \$100 bills without any training, received training on the design features of \$50 and \$20 bills, and then made judgments about \$50s and \$20s. Only offset counterfeits of \$50s and \$20s were used. This design allowed some comparisons of judgments of \$100 bills with and without training, and some comparisons of judgments of \$100, \$50, and \$20 bills. None of these comparisons could be done on all the judgment conditions due to constraints on available counterfeits. The goal of Experiment 2 was therefore to determine whether the results found for \$100s would also be found for \$50s and \$20s, and to see whether training would affect people's accuracy. Only global judgments were made with the \$100 bills; only offset counterfeits used were used with the \$20 and \$50 bills, and none of the trials were masked. These design choices were made based on the results obtained in the first experiment. With the smaller denomination bills, both featural and global judgments were made, but, unfortunately, it was not possible to obtain a sufficient number of counterfeit bills to complete the intended design (two replications of the 1990 vs. 1996 series, features within series, featural vs. global judgments, and denomination and two different sets of stimuli). This greatly limited the data analysis, as will be noted.

3.1 Method

In all respects other than those listed here, the design of Experiment 2 was the same as the design of Experiment 1.

3.1.1 Participants

The data to be presented are from a sample of 138 college student observers who received course credit for their participation. They had an average of 12.8 ± 1.1 years (range = 12 to 16) of schooling and an average age of 20.07 ± 3.9 years (range = 17 to 41). The sample was 74% female and was 1.5% Native American, 11.7% Asian, 16.8% African-American, 47.5% White, and 22.6% Hispanic. Answers to other demographic questions are in the Appendix.

3.1.2 Stimuli

The stimuli used for judgments of \$100s were the 2 sets of stimuli that had been created for global judgments of both series in Experiment 1 (Conditions FM90 and GU90). Approximately half the observers evaluated each set. The same features were of interest for \$50s and \$20s as for \$100s (6 features for 1996 series, 4 features for 1990 series), although the features are designed differently in the different denominations. The initial design of Experiment 2 called for the construction of two sets of 50 \$50 bill pairs and 50 \$20 bill pairs, using only unmasked offset counterfeits. However, lack of available counterfeits limited this phase of the study to one set of 62 pairs that all observers judged.

3.1.2.1 Featural Judgments

Two trials (and thus two stimulus sets) were created for each of the six features of the \$20s and \$50s in the 1996 series, resulting in twelve trials for each denomination. For \$20s in the 1990 series, no trials were created to test judgments about security thread, but two trials were created for each of the other three features, resulting in six trials. For \$50s in the 1990 series, two trials were created to evaluate microprinting and one trial was created to evaluate fine lines, resulting in three trials; portrait and security thread were entirely missing.

3.1.2.2 Global Judgments

Global judgments of 1996 series \$20s were complete: twelve bills matching those used in featural judgments plus six other bills in which at least one deterrence feature was missing from the counterfeit. Global judgments of 1990 series \$20s used six bills matching those used for featural judgments (note that security thread was again not evaluated) and four other bills that had at least one deterrence feature missing from the counterfeit. For \$50s, 2 bills matching those used to judge watermark in the 1996 series were used. No other \$50s from either series were used in global judgments.

3.1.3 Procedure

The session began with collection of demographic information. They were then given general instructions about the task, but no information about how to detect counterfeits. After 2 practice trials, the observer then either judged the 1996 series \$100 bills followed by the 1990 series \$100 or vice versa, in random order. Following this, the observer read “New designs for your money” published by the U.S. Department of Treasury describing the 1996 design of \$50s and \$20s, and the experimenter pointed out the features relevant to the study. The 62 \$20 and \$50 bill trials were then run, with counterbalancing of the four conditions (global judgments of \$50s, global judgments of \$20s, featural judgments of \$50s, featural judgments of \$20s) working as it did in Experiment 1.

3.2 Results

3.2.1 Overall accuracy and effect of experience

Over all trials, observers were 93.4% correct in judging \$100 bills, and 97.4% correct in judging \$20 and \$50. The trends relating accuracy with reported familiarity with counterfeiting deterrents were similar to those obtained previously in that those who had never heard of the features or had heard of them but not looked at them were less accurate than those who had looked at them but not tried to use them or had tried to use them, the group mean differences were not significant. The respective means were 92.0%, 92.4%, 95.4%, and 95.3% for \$100s and 97.0%, 96.8%, 98.1%, and 98.8% for \$20s/\$50s ($n = 33, 55, 34, \text{ and } 14$, respectively). Accuracy in judging \$100 bills was at least modestly correlated with accuracy in judging \$20 and \$50 bills ($r = .48$).

3.2.2 Effect of Training on Accuracy in Judging \$100s

An ANOVA was conducted on global judgments of \$100 bills in the two experiments to evaluate the effects of brief training. Whether or not the observers had received training (Experiment 1 vs. 2) was a between-subjects variable and series (1990 versus 1996) was a within-subjects variable. Judgments were 96.7% accurate with training and 93.3% accurate without training, a significant difference, $F(1, 208) = 12.12$. Judgments were 95.6% accurate for the 1996 series and 93.3% accurate for the 1990 series, also a significant difference, $F(1, 208) = 13.11$. There was no evidence that training produced different results for the two series, $F(1, 208) < 1$. Figure 9 describes these two effects graphically.

3.2.3 Accuracy in judging \$100 bills as a function of the characteristics of the currency

Table 4 contains the mean accuracies obtained in judging the \$100 bills in this experiment without any training. The data were evaluated using a series of ANOVAs. The first examined performance in judging 1996 series bills as a function of inkjet vs. offset counterfeiting processes and the seven deterrence features; the second examined performance in judging 1990 series bills as a function of counterfeiting processes and the six deterrence features, and the third compared the two series, the two processes and the six deterrence features common to both. The only significant effect within the 1996 series was the accuracy difference between inkjets and offsets processes (96.4% vs. 92.2%), $F(1, 137) = 44.19$. Similarly, processes was the only significant effect within the 1990 series (95.7% vs. 88.5%), $F(1, 137) = 36.79$. Figure 10 contains these two corresponding effects.

3.2.4 Accuracy in judging \$20 and \$50 bills as a function of the characteristics of the currency

Because of the various missing cells for the \$20 and \$50 bills, analyses had to proceed obliquely. First, a series of

simple ANOVAs were conducted examining feature differences for each possible combination of 1990 vs. 1996 series, featural vs. global judgment, and \$20 vs. \$50 bills. This could not be done with global judgments of \$50 in either old or new series because there were not at least two features to compare. Table 5 contains the mean accuracy of judgments of \$20s and \$50s for each available feature, judgment, and series. First, the differences among features were significant with featural judgments of the 1996 series of \$50 bills, $F(5,685) = 3.64$. They were likewise significant with featural judgments of the 1996 series of \$20 bills and with featural judgments of the 1990 series of \$20 bills, $F(5,685) = 8.30$ and $F(2, 274) = 5.63$, respectively. The remaining comparison of featural judgments of the 1990 series of \$50 bills approached, but did not achieve the designated significance level, $F(1,137) = 4.14$. Figures 11-14 depict these trends as a function of feature graphically.

Next, a series of ANOVAs were run examining differences between denominations and/or series. First, differences among the six features common to featural judgments of \$20 and \$50 bills were compared within the 1996 series. The main effect of feature, the main effect of denomination, and their interaction were all significant, $F(5,685) = 5.80$, $F(1,137) = 8.87$, and $F(5,685) = 7.57$. Watermarks were judged most poorly; fine lines were judged somewhat more accurately, and the remaining features were then all judged with nearly equal accuracy. The \$50 bills were judged more accurately than \$20 bills. However, the major reason both main effects arose was that watermarks on \$20 bills were judged anomalously low compared to the remaining combinations of features and denominations.

Next, differences between featural and global judgments for \$20 1996 series bills were evaluated over the six features common to them. The main effects of feature and type of judgment were all significant, $F(5,685) = 7.16$, $F(1,137) = 19.25$, as was the interaction between them, $F(5,685) = 6.24$. These results arose because there were difference across features when featural judgments were made, but not when global judgments were made, again an indication of the successful overall matching of the bills. However, no significant differences were found in a parallel analysis of \$20 bills from the 1990 series. Finally, differences between the two features common to featural judgments of \$20 and \$50 bills within the 1990 and 1996 series (fine lines and microprinting) were compared. Four effects were significant: the main effect of denomination, the main effect of series, their interaction, and the interaction between feature and denomination: $F(1,25) = 31.92, 29.53, 8.10$, and 24.51 . The first three effects arose because \$20 bills from both series and \$50 bills from the 1996 series were judged with fairly high accuracy (96.6% to 97.3%, but \$50 bills from the 1990

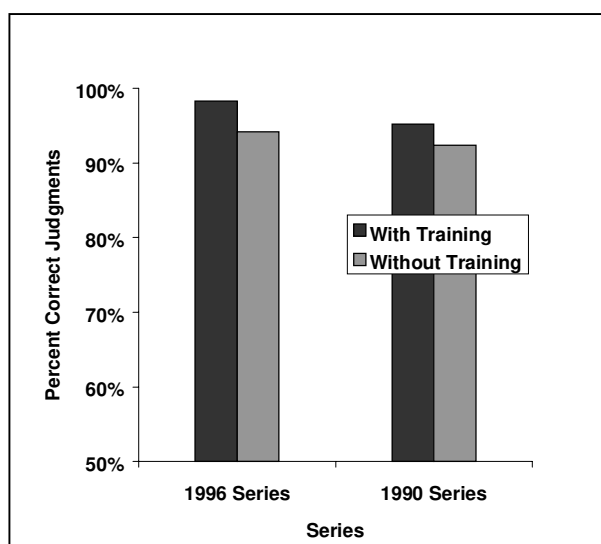


Fig. 9: Effect of Training and Series on Judgments of \$100 bills (Both experiments)

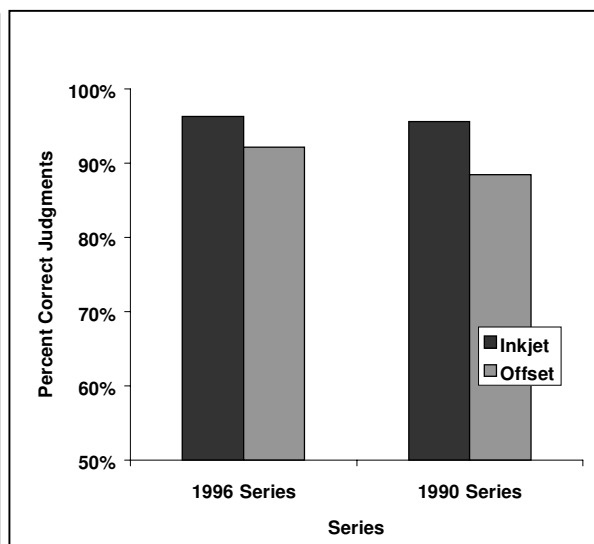


Fig. 10: Effect of Series and Process on Judgments of \$100 bills (Experiment 2)

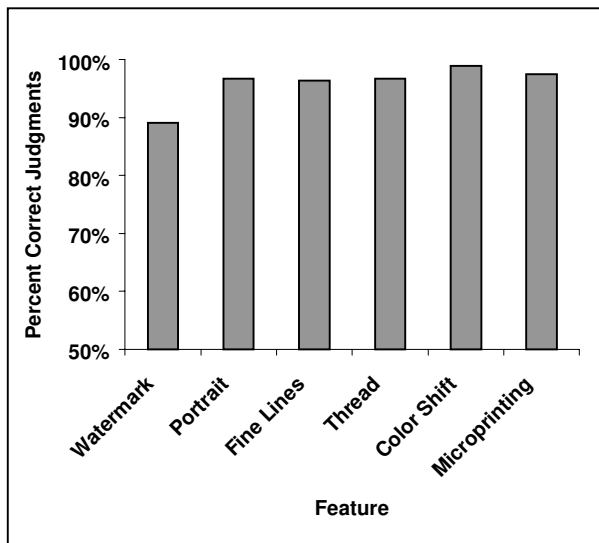


Fig. 11: Effect of Feature on Featural Judgments of \$20 bills-1996 Series (Experiment 2)

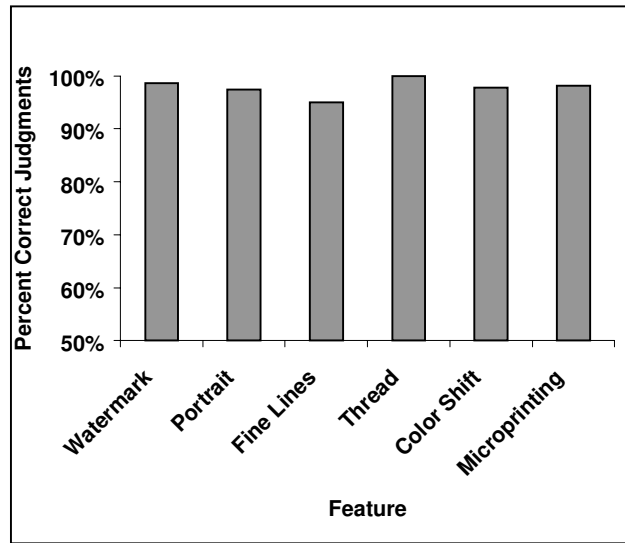


Fig. 12: Effect of Feature on Featural Judgments of \$50 bills-1996 Series (Experiment 2)

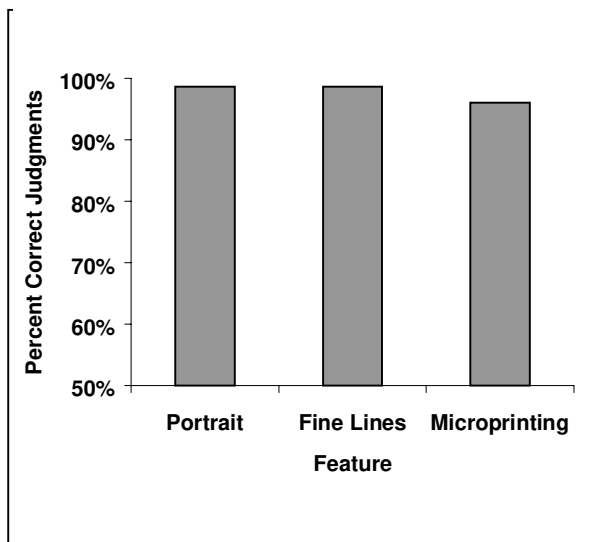


Fig. 13: Effect of Feature on Featural Judgments of \$20 bills-1990 Series (Experiment 2)

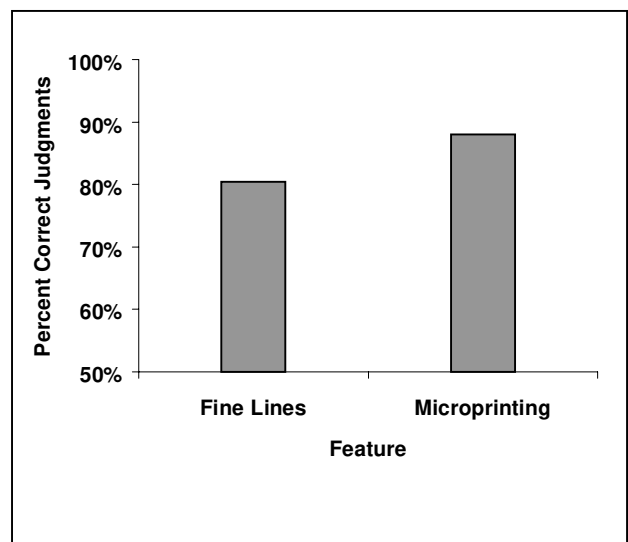


Fig. 14: Effect of Feature on Featural Judgments of \$50 bills-1990 Series (Experiment 2)

Series	Process	Watermark	Portrait	Fine Lines	Thread	Color Shift	Microprinting	Missing
1996	Inkjet	96.7	97.5	94.9	97.1	96.7	96.0	95.7
	Offset	93.1	93.5	87.7	94.6	92.8	91.7	92.1
1990	Inkjet		97.1	97.1	90.2		96.4	97.5
	Offset		89.1	83.3	91.3		88.0	90.6

Table 4: Percentage Correct in Judging \$100 bills in Experiment 2 as a Function of Series, and Feature

Denom.	Series	Judgment.	Watermark	Portrait	Fine Lines	Thread	Color Shift	Microprinting	Missing
\$50	1996	Featural	98.6	97.5	94.9	100.0	97.8	98.2	
\$20	1996	Featural	89.1	96.7	96.4	96.7	98.9	97.5	
\$50	1996	Global	99.3						
\$20	1996	Global	98.6	99.3	99.6	98.9	98.9	98.6	99.2
\$50	1990	Featural			80.4			88.0	
\$20	1990	Featural		98.6	98.6			96.0	
\$20	1990	Global		98.6	98.6			100.0	99.1

Table 5: Percentage Correct in Judging \$20 and \$50 Bills as a Function of Denomination (Denom), Judgment Condition, Currency Series, and Feature

series were judged much less accurately (84.2%). The remaining interaction arose because fine lines and microprintings were judged with nearly equal accuracy in \$20s (97.4% and 96.7%), but fine lines were judged more poorly than microprintings with \$50 bills. This effect is only included for completeness; it is of minimal present interest because it is common to both 1996 and 1990 series bills.

3.3 Discussion of Experiment 2

There were at most small differences in people's ability to detect counterfeits in the different denominations. People were better at detecting counterfeit \$20s, and we speculate that may be because of their greater familiarity with that denomination.

People were better at detecting counterfeits after having the deterrence features explained to them than they were with no training. People were also better when they reported having familiarity with the deterrence features before the experiment than when they reported no previous experience with them. But even in the group with the least experience with the deterrence features, detection was surprisingly good.

In sum, our current results with \$100 bills are consistent with those previously observed. Likewise, we can make no statements about differences among features of \$20 and \$50 bills save to say: (a) there is no glaring discrepancy between what previously was found, and (b) perhaps there is a role played by observers' greater familiarity with \$20 bills. Perhaps the most interesting finding is the difference in accuracy of judging \$100 offset bills as a function of what is in effect minimal training in counterfeit deterrence, as seen in the overall means in the previous experiment vs. the first part of this experiment. Finally, an apparent decrement in judging \$100 bills using fine lines alone that was observed in Experiment 1 did not replicate here.

4. GENERAL DISCUSSION

Our results indicate that observers whose attention is focused on discriminating genuine bills from counterfeits perform the task quite well except in the most difficult condition (masked bills, featural judgments on 1990 series, offset-printed, counterfeit \$100 bills). Under the best of circumstances (training, availability of the whole bill, use of inkjet counterfeits) subjects were very accurate even with the 1990 series bills, making it difficult to document improvement with the 1996 series. But despite this difficulty, participants were, in fact, more accurate in judging the 1996 series than the 1990 series. Unfortunately, it is difficult to assess the improvement provided by the modified features because of the differences in their baselines and the ceiling effects present, especially with inkjet counterfeits. In other words, judging the improvement of features is not the same as judging their absolute efficacy. Perhaps the best thing to say is that there are only a few cases involving the 1996 series warranting attention—the watermark used on \$20 bills and the watermarks, portraits, fine lines, and microprinting used on the \$100 bills. The poor performance of the last-named four improved under the more realistic unmasked conditions, and only the portraits and microprinting warrant further attention.

That said, one of the primary goals of this research was to see if the design changes made in the 1996 series improved observers' ability to detect counterfeits. This was assessed for \$100s, \$50s, and \$20s in different experiments. Although using different contexts limits one's confidence in comparing results for \$100s to results for \$50s and \$20s, it appears that the features are not necessarily equally effective in the different denominations in the 1996 series. Accuracy was not poor for any feature in any series, but watermarks were less effective for \$20s than for \$50s and \$100s, portraits and microprinting were less effective for \$100s than for \$50s and \$20s, and color-shifting ink may have been less effective for \$100s than for \$50s and \$20s. Although the 1996 series improved the effectiveness of portraits and microprinting in the \$100s, it is interesting to note that other results show that there is room for more improvement.

This experiment did not speak to the reason that genuine 1996 series bills were less confusable with counterfeits than 1990 series bills were. In particular it is unknown to what extent the improvements are due to the lesser amount of time that counterfeiters have had to practice and learn how to counterfeit the newer series. Counterfeiters could simply have had more practice with the older series and their "product" may improve with time. This can eventually attenuate differences between counterfeits the two series. On the other hand, as the 1996 series becomes more familiar to the

public, detection skill might increase. Both studies found a small benefit to detection of having had previous exposure to the counterfeit deterrence features. A longitudinal study would best determine whether time helped counterfeiters or detectors more.

Finally, it is important to return to the fact that people are much better at detecting counterfeits when focused on doing so, as they were in this experiment, than when focused on monetary transactions. The design changes clearly improved the degree to which counterfeits can be detected, but previous research has shown that in order for detection to occur, money handlers must learn about and pay attention to the counterfeit deterrence features on their bills.

ACKNOWLEDGMENTS

We thank Carson Clanton, Jennifer Gerber, Lisa Itgen, Umma Khan, Lorri Lancashire, Lori Lillie, Kimberly McConnell, Carmen Nephew, Richard Schell and Amy Waldrip for their assistance in data gathering. We also thank the US Secret Service for supplying counterfeit currency. The Bureau of Engraving and Printing of the Department of the Treasury, contract number TEP-00-07(N), supported this study. However, the conclusions reached herein are those of the authors and not the Department of the Treasury or any of its subdivisions.

REFERENCE

1. B. L. Collins, S. E. Mayerson, and J. A. Worthey, *NBSIR 85-3160: Noticeability of features of secure documents*. National Bureau of Standards (now NIST), Washington, D.C., 1985.

APPENDIX

Responses to Questions about Money Handling Experience

- Q1: For how many years have you worked with money in a retail setting?
Median responses: Experiment 1: 3-5 years; Experiment 2: 1-3 years
Modal responses: Experiment 1: 3-5 years; Experiment 2: 0-1 year
- Q2: For how many years have you used US money for your own purchases?
Median and modal responses, both experiments: 10-20 years
- Q3: How many years ago were you first trained on counterfeit detection?
Median and modal responses, both experiments: Never.
- Q4: What was your familiarity with the new counterfeit deterrence features before you were contacted about this study? Median and modal responses, both experiments:
I had heard what the features are but I had not looked at them.
- Q4: How frequently do you handle \$100 bills?
Median and modal responses, Experiment 1: Less than one a week
- Q5: How frequently do you handle \$20s, \$50s, and \$100s?
Median response, Experiment 2: 1-3 each day
Modal response, Experiment 2: Less than one a week